我国数据故事化的研究进展:内涵、流程、模型与应用[1]

刘桂锋 吴雅琪 刘琼

江苏大学科技信息研究所 镇江 212013

摘要：[目的/意义]数据故事化是一种更高效、更全面、更新颖地展示数据分析结果的方式通过分析我国数据故事化研究的代表性文献，对现有研究成果进行系统梳理，并对研究势趋进一步展望，为后续研究提供借鉴和启示。[方法/过程]主要采用内容分析法归纳我国现有的有关数据故事化研究的21篇文献内容，从概念内涵、基本流程、相关模型以及应用实践4个方面对我国数据故事化的研究进展进行梳理和分析。[结果/结论]数据故事化研究在我国尚处于起步阶段，在内涵、流程、模型、方法、工具等理论研究以及实践应用领域存在较大提升空间，未来研究注重构建我国数据故事化研究的自主知识体系和应用实践体系。

关键词：数据叙事；数据故事化；数据科学；科学数据；数据管理

分类号：G254

1 引言

2023年2月28日中共中央办公厅印发了《数字中国建设总体规划》，大力推进数字中国建设，提出数字中国建设是数字时代中国现代化建设的重要引擎，是打造国家新竞争力的强大支柱。数据作为一种新型的生产要素，充分利用数据资源，是推进数字中国建设的基础。然而，现有的数据存在着复杂、混乱、枯燥的问题，公众难以获取有价值的数据资源，更难以揭示数据之间的关系，所以如何从海量的数据中获取价值，并将这些价值传递给人，成为了一个必然需求[1]。数据故事化（Data Storytelling）作为大数据时代的一项新兴课题，可以用来揭示数据之间的隐藏关系，洞察数据的隐藏价值，为数据发声、解释、说服和引导，提高公众对数据的接受度。

目前，国外关于Data Storytelling的理论和实践研究都取得了一定进展，如Dykes[2]等提出有效的数据故事化需要通过数据、叙事和视觉效果来驱动；Lee[3]等提出了发现见解、将这些见解转化为故事以及将这种故事传达给观众的数据故事化流程； McDowell[4]等描述了

[1] 本文系国家社科基金一般项目“科学数据融合模式设计与体系建构研究”（项目编号：21BTQ080）的研究成果。

作者简介：刘桂锋(ORCID: 0000-0002-7209-3862)，研究馆员，硕士生导师，博士，liuguifeng29@163.com；吴雅琪，硕士研究生；刘琼(ORCID：0000-0003-0215-556X)，馆员，硕士。

故事化如何扩展数据、信息、知识和 DIKW 金字塔并修订了一个新的 S-DIKW 框架；Brolcháin[5]等分析了现有的开放数据平台对故事化的支持，认为 YDS 平台可以用作为现有开放数据平台开发相关故事扩展的模板；Ojo[6]等分析了 GEN 数据新闻奖的 44 个获奖案例，认为数据故事化可以应用于数据新闻的传播；Botsis[7]等提出数据故事化可以应用于增强生物医学科学的知识传播。我国学者在国外研究的带动下也对数据故事化展开了一定的探索，主要集中在理论分析与实践应用等方面的模块化分析，暂时还没有进行系统的综述，缺乏全面的研究进展总结。因此，本文在详细梳理我国现有的 21 篇有关数据故事化研究论文的基础上，归纳总结“数据故事化”的概念内涵、基本流程、相关模型以及实践应用（见图 1），以对我国数据故事化的研究进展进行全面的剖析。

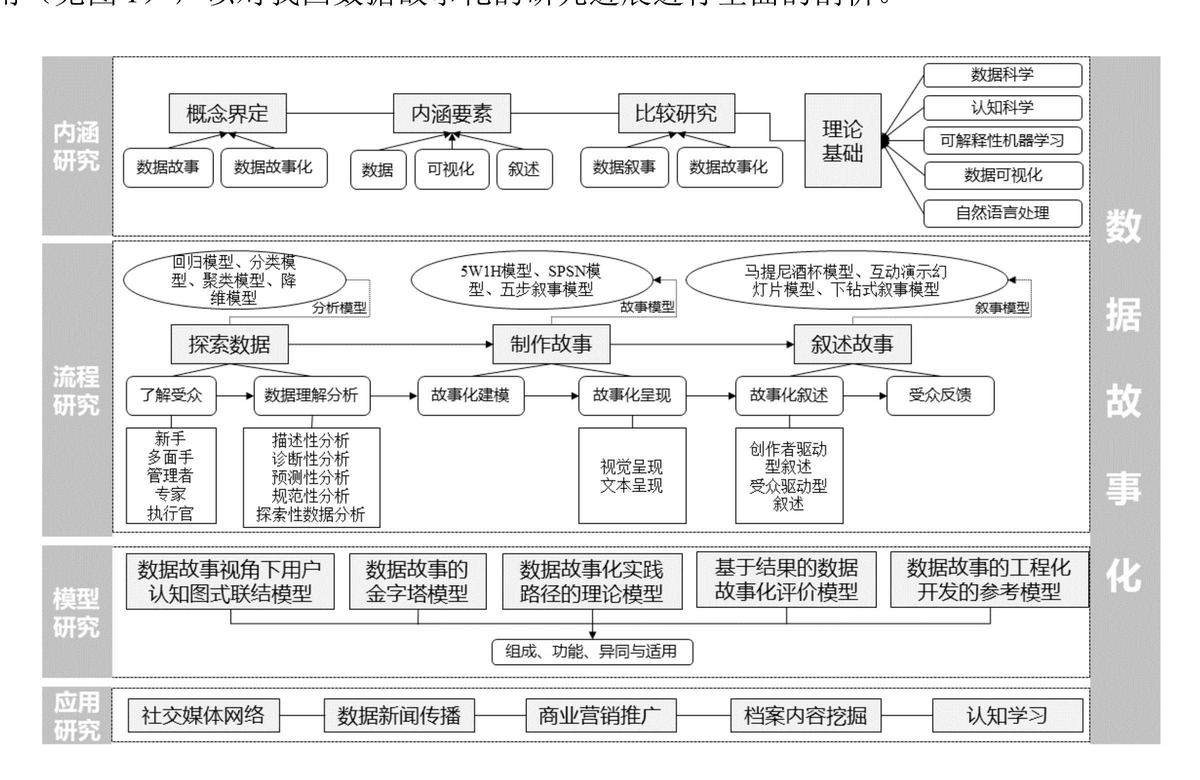


图 1 数据故事化的研究框架

2 数据故事化的内涵研究

2.1 概念界定

目前国内学者对于数据故事化的概念界定各有侧重，本文通过对现有文献分析后，认为数据故事（Data Story）是一种数据驱动的故事形式，解决了特定的业务需求，使用数据作为输入，应用数据分析和建模技术，从数据中提取有价值的见解，并以故事的形式传达给目标受众[8,12]。而数据故事化（Data Storytelling）本质上就是从数据到数据故事的

一个转化过程，是以“故事”的形式呈现对“数据的理解”。在数据故事化中，“数据”被缩小或与特定情况相关，并以叙述的形式呈现，使其更容易理解、记忆和消化[12,18]。数据故事是数据故事化的最终形态，是一种文学体裁，其中数据是叙事的主体，数据的客观性和叙事的主观性形成一个整体。对许多人来说，关于数据的故事比数据本身更容易被记住、感知和体现出来。

2.2 内涵要素

根据朝乐门[13]的研究，数据故事化主要涵盖了三个关键因素：数据、可视化和叙述（见图 2）。在数据故事化的过程中经常会用到数据可视化技术，但重要的是不只是将图形呈现出来，而是逐步引导人们认识和理解数据，并从数据中得出结论。将数据与可视化相结合，可以提供数据的视觉表现，让其他人更好地理解数据并从数据集中得到更多的认识。当在解释数据的过程中应用叙述时，可以让人们更好地了解到数据中发生的事情，并理解其背后隐藏的信息。当把可视化与叙述结合在一起时，可以吸引他人，以一种更容易理解和接受的方式来吸引受众。而当有价值的数据、有效的可视化和熟练的讲故事方式结合在一起时，一个关于数据的引人入胜的故事就产生了。

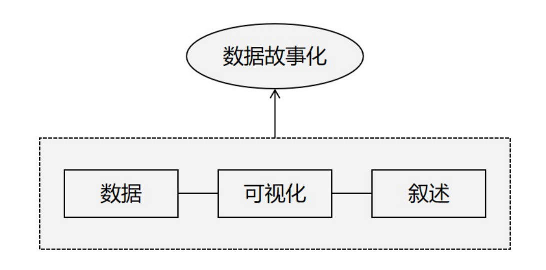


图 2 数据故事化三要素

2.3 数据故事化与数据叙事的比较研究

通过分析现有文献发现我国学者主要把 Data Storytelling 理解为“数据故事化”与“数据叙事”两个内涵角度。本文整理归纳了一些代表性的相关研究（见表 1），认为“数据故事化”与 “数据叙事”并列，是一种呈现数据的方式。数据故事化更侧重于关注新技术环境下的组织和数据表达创新，并提供了一条通过可视化将数据意识深化为数据认知的途径；而数据叙事则更注重分析数据中的事件、背景和人物等元素，以描述数据之间的隐性关系并传达相关信息。但这两个概念在本质上是相同的，其基本思想是利用某些技术方法和策略来重构数据的故事化。

表 1 数据故事化与数据叙事

Data Storytelling	研究人员	研究内容	侧重点
数据故事化	王萍[19]	研究了美国的开放数据平台案例并提出了政府开放数据故事化的运行模式。	侧重于新技术环境下的数据构成和呈现层面的创新。
	孙智中[20]	研究了数据故事化的评价要素并提出了基于结果的数据故事化评价模型。	
	周霞[21]	研究了数据故事化在数据新闻中的应用案例并提出了数据故事化实践路径的理论模型。	
	赵红颖[14]	研究了数据故事化是如何讲好红色档案故事的。	
	候雪林[16]	研究了疫情数据故事化的要素结构、系统结构以及层级结构并提出了疫情数据故事化的实现过程。	
	孟刚[18]	研究了数据故事化中的数据关键特征并提出在关键特征视角下的效率和质量评价。	
数据叙事	张斌[22]	研究了一些数据叙事驱动的馆藏利用模型并提出了其基本架构。	侧重于如何利用叙事策略对数据中的事件、背景、人物和其他元素进行分解和连接。
	王海智[2323]	研究了数据叙事在解决数据新闻现存问题的路径。	
	张晨[9]	研究了支持数据叙事的关键技术。	
	丁家友[24]	研究了数据叙事的基本要素与不同的叙事策略可以应用的场景。	
	陈昱彤[17]	基于 5W 传播模式基础提出了数据叙事的运行模型并研究了其运行中的关键问题。	
	李锦绣[25]	研究了数据叙事的应用情境与实现路径。	
基本思想均在于通过一定的技术方法及策略完成数据的故事化重塑			

2.4 理论基础

数据故事化的研究需要对其他相关领域的研究基础进行借鉴，本文通过总结归纳现有文献，认为数据故事化的理论基础主要涉及数据科学、认知科学、数据可视化、可解释性机器学习和自然语言处理五个领域（见表 2），表中简要阐述了其他各个领域研究与数据故事化之间的联系，分析其是如何支撑数据故事化的研究，并对其之间的联系进行了图示展示。

表 2 数据故事化的理论基础

理论基础	联系	图示	参考文献

数据科学	数据科学专注于数据处理、计算、管理、分析和数据产品开发等关键活动，为数据故事化的过程提供了启示与技术背景，数据故事化从学科归属上来看是属于数据科学的一部分。	数据故事化 数据科学	丁家友，2022[24]
认知科学	认知科学关注的是信息如何以感觉语言、注意力、推理和情感的形式呈现，数据故事化认知过程的每个阶段都需要认知科学的理论指导。	数据故事化的认知过程 认知科学 感知 认知 行动	张晨，2019[13]
可解释性机器学习	可解释性机器学习是指以人类、特别是非专业人士能够理解的方式来表示数据的能力，为数据故事的建模提供了重要的指导方针，为理解算法和解释数据故事中的结果提供了一个理论框架。	故事创作者 分析模型 故事化建模 故事模型 可解释性机器学习	肖纪文，2023[12]
数据可视化	数据可视化通过一些可视的方式来呈现数据，是数据故事化中最常用的叙事工具之一，可视化技术可以提高对数据故事的理解力。	故事叙述者 故事模型 故事化呈现 故事叙述 数据可视化	陈昱彤，2022[17]
自然语言处理	自然语言处理的主要研究目标是使计算机能够理解、解释和处理人类语言，其中自然语言理解将输出数据处理成机器可读的语义表示，是数据故事化建模的关键技术之一；自然语言生成以人类可读的自然语言形式表示语义信息，是数据故事化呈现的关键技术之一。	自然语言处理 故事化建模 自然语言理解 自然语言生成 故事化呈现	朝乐门，2021[26]

3 数据故事化的流程研究

数据故事化本质上属于一种传播信息的方式，它以一种更易于受众记忆、认知和体验的方式将数据应用到故事化的整个流程中。基于现有的文献[11131517192126]，本文将数据故事化的基本流程总结归纳为六个基本活动：了解受众、数据理解分析、故事化建模、故事化呈现、故事化叙述、以及受众反馈，也可将其分为探索数据、制作故事和叙述故事三大部分（见图3）。

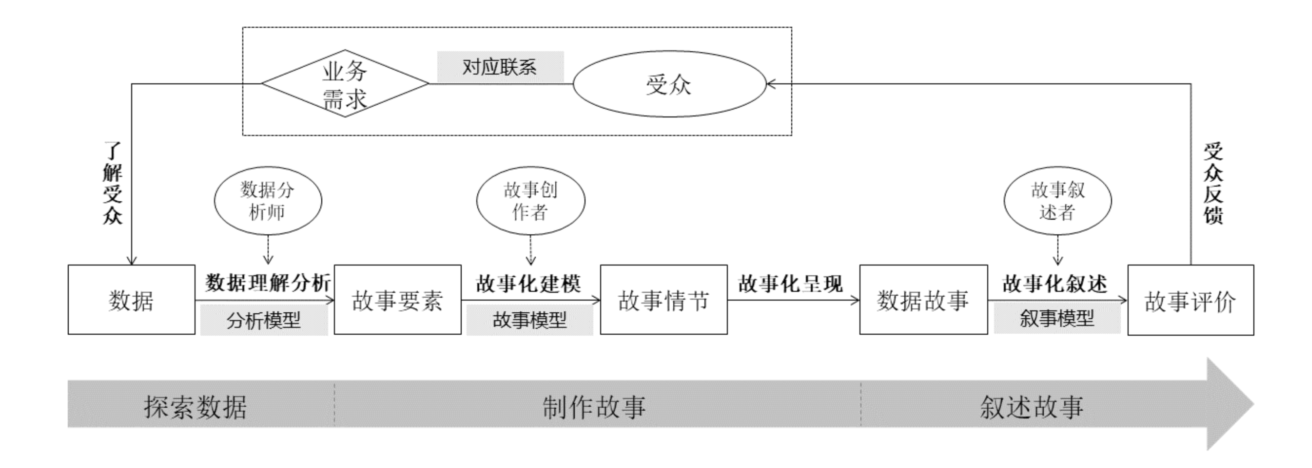


图 3 数据故事化的流程研究

3.1 探索数据

数据探索涉及一系列旨在理解和分析数据的活动。数据是原材料，关于数据的视觉故事的内容就是从这里产生的。在探索数据的过程中，首先要细分受众群体，了解不同受众群体的不同业务需求，再由数据分析师根据受众的业务需求对数据进行理解分析，以达到洞见数据隐藏价值的目的。

3.1.1 了解受众

了解受众群体是数据故事化的先决条件。孟刚等[18]在考虑受众对故事的了解程度和信息获取能力的基础上将受众划分为五种群体：新手、多面手、管理者、专家和执行官。新手通常是刚接触这个话题，希望数据故事可以更详细地展示出来，为新手提供更详细和更具体的信息。多面手通常对当前题有了一定的认知，但寻求一个概述和关键问题，更多的是需要关注总体理解和核心主题。管理者通常寻求对话题细节的错综复杂和关系的深入和实际理解。专家通常希望对所研究的主题有更多的了解，而不需要深入到不必要的细节。执行官通常只需要知道概率计算的结论及其影响，主要关注重要性和结论。

因此，根据受众的范围、知识、技能和目标，数据故事专家必须确定故事的适当背景。这意味着要使不同的数据故事适应不同的数据故事业务需求，对每个受众进行细分，了解受众的需求，将受众群体与业务需求之间进行对应联系，根据不同的受众选择不同的数据故事和内容呈现方式，不要过度泛化，要使受众能够准确理解并感受到数据的价值（见表 3）。朝乐门[26]等提出业务需求主要包括建议、描述、解释、劝说、调查、研究、教育 7 种类型。建议是向受众推荐产品或服务的过程。描述是向受众解释数据或信息的过程。解释是向受众解释想法和结论的过程。劝说是说服听众接受某种想法或观念的过程。调查是向受

众收集用户数据和意见的过程。研究是通过受众的参与来了解商业和产品创新与优化的过程。教育是让受众掌握知识和技能的过程。

表 3 受众群体与业务需求之间的对应联系

受众群体	需求内容	业务需求种类
新手	寻求数据故事能够详细展示，希望提供更详细和具体的信息。	解释/推荐
多面手	寻求一个概述和关键问题，需要关注总体理解和核心主题。	描述/研究
管理者	寻求对话题细节的错综复杂和关系的深入和实际理解。	劝说/研究
专家	寻求对所研究的主题有更多的了解而不需要深入到不必要的细节。	调查/研究
执行官	寻求知道概率计算的结论和其影响主要关注重要性和结论。	劝说/教育

3.1.2 数据理解分析

数据理解实际上是进行数据故事化的第一步，它本质上就是对数据进行分析。数据故事与文学故事的一个主要区别也是数据理解。本文通过分析现有的文献[13162627]，发现在数据故事化中常用于数据分析的方法有四种: 描述性分析（Descriptive Analytic）、诊断性分析（Descriptive Analytic）、预测性分析（Descriptive Analytic）和规范性分析（Descriptive Analytic）四种（见图 4）。

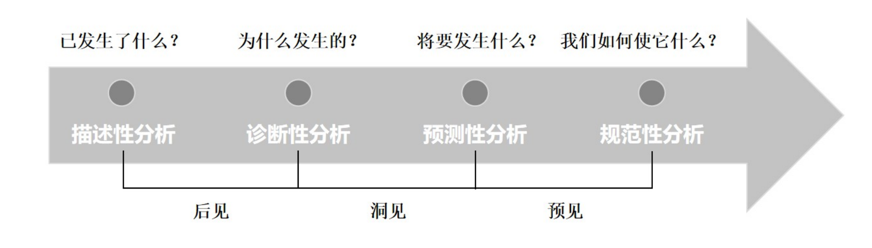


图 4 数据分析的四种方法

描述性分析侧重于“过去”，作为数据分析的第一步，回答“发生了什么”的问题。诊断性分析侧重于“过去”，回答 “为什么会发生”的问题，作为对描述性分析的进一步理解。预测性分析关注“未来”，回答“会发生什么”的问题，作为规范性分析的基础。 诊断性分析侧重于“过去”，并回答“为什么会发生”的问题，这是描述性分析的另一种理解。 预测性分析侧重于“未来”，回答“会发生什么”的问题，是规范性分析的基础。在进行上述四种数据分析的基础上，需要进一步进行探索性数据分析（Exploratory Data Analysis，EDA），以便达到更好地理解数据的目的。探索性数据分析指的是旨在识

别基础数据中的模式和趋势的数据分析活动。

在数据理解分析的过程中，数据分析师还需要运用分析模型来对业务需求进行描述。运用分析模型就是将数据与业务需求相结合的过程，即为了将数据转化为故事而进行的统计分析，包括了许多机器学习模型，通常有分类模型、聚类模型、回归模型、关联规则模型和降维模型等[9202226]。

3.2 制作故事

制作故事涉及以故事化建模与故事化呈现为中心的一系列活动。创作者需要将故事放在受众的业务领域、知识、技能和目标的背景下，根据其定制各种数据故事，使其易于理解。为了制作一个故事，数据故事专家需要将第一步中数据分析师提取出来的故事要素进行故事化建模形成故事模型，然后通过形式化描述将其组合成一个有趣、富有启发性和引人入胜的故事情节，再通过一定的可视化技术呈现出来，生成可以叙述传播的数据故事。

3.2.1 故事化建模

从数据到故事模型的转化是数据故事化的关键活动[12]。故事模型是故事创造者和叙述者之间的桥梁。在了解受众以及理解和分析数据等活动之后，需要对数据故事进行建模（见图 5）。数据故事建模的一个重要部分是确定数据故事的组成部分，建立并验证各元素之间的内部关系，并确定不同故事之间的关系，这包括确定数据故事的要素，即数据故事的七个基本要素（人物、需求、情节、情境、冲突、解决方案和下一步行动）；建立并验证数据故事各要素之间的内部关系，即其他以冲突为基础，组织数据故事的六个要素，分析内部关系，验证各要素之间的一致性，并建立不同故事之间的联系，并将其联结起来，再代入故事模型就可以形成一个完整的故事。

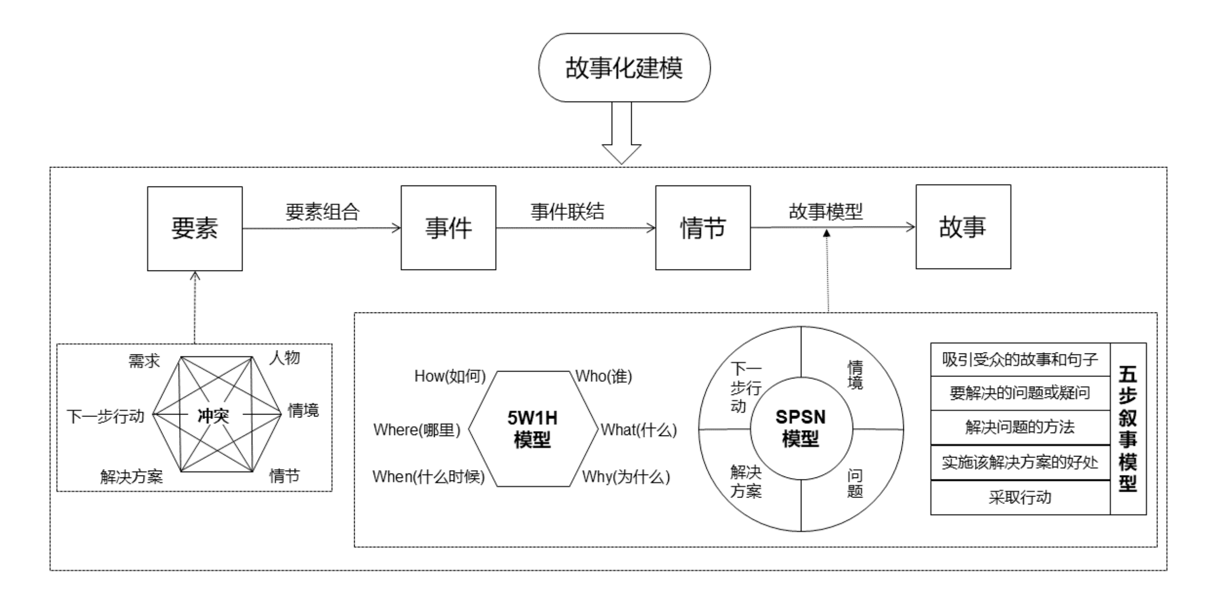


图 5 故事化建模过程

故事模型主要描述的故事要素及要素之间的结构关系[112627]，比较有代表性的如下：（1）5W1H 模型假设数据故事必须包含六个元素：Who（谁）、What（什么）、Why（为什么）、When（什么时候）、Where（哪里）和 How（如何）。Who（谁）对应数据故事的叙述者，What（什么）对应数据故事的具体内容，Why（为什么）对应讲述数据故事的原因，When（什么时候）对应数据故事发生的时间，Where（哪里）对应数据故事发生的地点，How（如何）对应数据故事是如何发生的。（2）SPSN 模型描述了故事的要素和结构模型，认为数据故事包括的要素及其描述顺序依次为：Situation（情境）、Problem（问题）、Solution（解决方案）、Next Steps（下一步行动）。情境是指数据故事叙述者向听众描述故事发生的情况以及他/她想要改变的初始状态；问题指的是解决讲故事者最初情况下的问题或痛点；解决方案指的是提出的解决方案；下一步指的是数据叙述者和受众要采取的行动。（3）五步叙事模型指出数据故事的基本要素和结构主要包括五个内容：吸引受众的故事和句子，要解决的问题或疑问，提出解决问题的方法，描述实施该解决方案的具体好处，并呼吁采取行动。

3.2.2 故事化呈现

从故事模型到故事的叙述也是数据故事化的关键活动[9]。张晨[9]等提出故事化的呈现方式主要分为两种类型，即视觉呈现和文本呈现。视觉呈现强调基于数据可视化的数据故事化的重要性，一般数据占较大比例，是一种视觉驱动的方式，主要应用于分析、探索和解释；而文本呈现则侧重于利用非可视化的实现方式将数据置于特定的叙事背景中，具有更强的文学性，一般数据占比比较小，是一种听觉驱动的方式，主要应用于描述和解释。

基于现有文献[8131826]，笔者本文总结归纳出支持视觉呈现的关键技术包括注释、图表自动生成、可视化等技术，支持文本呈现的包括自然语言生成和文本转换等技术。结合视觉和文本两种方式来创造和呈现故事产品的技术包括富媒体、人机互动、虚拟现实和增强现实等技术（见图 6）。

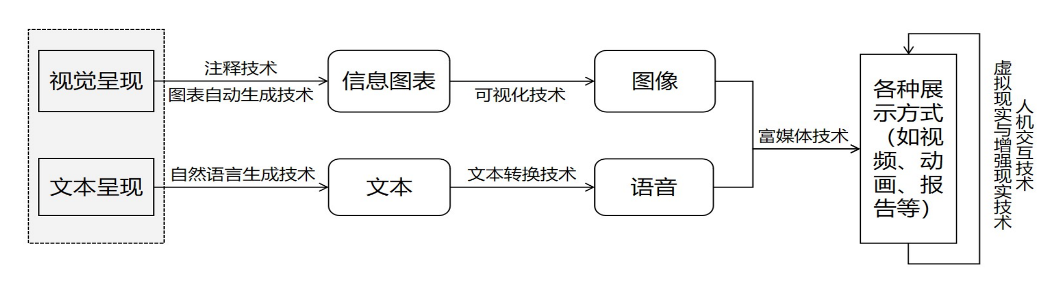


图 6 支持故事化呈现的关键技术

在视觉呈现的过程中，通常会创建出一个信息图表或是生成一个图像。在这个过程中，想要创建出信息图表的关键就是要选择正确的图表来表达数据，而对特定图表的选择取决于故事创作者的目标和数据的类型。利用图表自动生成技术可以输出自定义或自动推荐的视觉图表类型，然后根据用户输入的数据创建信息图表。注释技术可以用来为信息图表添加图形或文字标签，以直观地突出重要的特征或数据元素，并促进对可视化结果的特定区域的理解和解释。此外，还可以利用可视化技术，将它用于特定的界面中以来表达可视化的元素，然后允许用户通过选择、改变、调用和转移可视化界面元素来改变界面表示结果，并提供不同的图像生成技术的视角。

在文本呈现的过程中，通常会形成一些文本或者语音。在这个过程中，可以利用自然语言生成技术将一些晦涩难懂的数据故事模型生成一段人类可理解的自然语言文本或者对数据故事进行补充描述。此外，还可以利用文本转换技术对生成的文本进行语言分析使其生成一段语音，其本质上就是将文本输入转化成语言输出的一种自动化技术。

数据故事化的依托载体和生成结果为数据故事产品，因此数据故事化的过程也是开发和使用数据故事会产品的过程。可以利用富媒体技术、人机交互技术、虚拟现实或增强现实技术将视觉呈现与文本呈现结合在一起，形成的最终产品可以是动态或静态、单页叙述或互动研究的形式，例如，图像、报告、音频、视频、动画等。富媒体技术是指使用各种技术，主要是图像、动画、音频、视频、脚本代码和标记语言，来创造各种广告产品、电影产品或网络内容，让用户可以与之互动。基于人机互动的数据故事，涉及受众和计算机之间的互动，以完成一个学习过程，计算机可以收集用户数据并提供适当的反馈。虚拟现实技术可以通过赋予数据场景新的虚构现实，让受众沉浸其中，从而传达出数据场景的本质。增强现实技术可以对原始数据场景进行优化和调整，将其置于一个真实的环境中，以 3D 方式讲述故事，通过让观众沉浸在场景中的能力来吸引他们，并为数据产品创造一个更好的体验。

3.3 叙述故事

叙述故事涉及以故事化叙述以及受众反馈为中心的一系列活动。首先故事叙述者可以用各种讲故事的方式向受众叙述传播数据故事，受众再根据自己感知对数据故事进行评价，最后将受众的评价反馈给创作者。

3.3.1 故事化叙述

从故事的叙述者与受众之间的驱动关系看，数据故事的叙述者可采用的叙述方法有叙述者驱动型叙述方法和受众驱动型叙述方法。（1）叙述者驱动型叙述的重点是向观众

提供数据，叙述者在讲故事的过程中主要是走直线，不给观众与图表的互动，数据和可视化的内容由叙述者选择，并作为成品呈现给读者与观众，在这个过程中，受众一般都是被动接受的，这是一种片面的展示、交流和获得接受的方法。这种技巧经常用于叙述者作为主要演员的活动中，如电影、书籍和杂志、广告和商业演示，其目的是为了有效沟通，让他人接受确定性的信息。（2）受众驱动型叙述强调与观众的高度接触，并为读者提供处理数据的机会。讲故事的人负责提供数据及其可视化方法，而不是简单地告知观众并依靠严格的结构化叙事结构，观众通过图形化的可视化结构和形成他们自己的叙事流程来参与到这个过程。在这个过程中，受众一般都是主动的接受。数据故事的叙述需要在讲故事的人主导的叙述和受众主导的叙述之间取得平衡。讲故事的人向听众提供结构化的叙述，但互动仍然是可能的。

在故事叙述者向受众叙述故事的过程中还应用到了叙事模型，叙事模型是用来向目标受众讲述故事的模型[12132426]。同一故事模型可以存在多个叙事模型，以便使叙事个性化。在进行叙述者驱动型叙述时可以运用例如马提尼酒杯模型和 SUCCESs 模型等叙事模型，在进行受众驱动型叙述时可以运用例如互动演示幻灯片模型和下钻式叙事模型等叙事模型（见表 4）。叙事需要一个讲故事的过程，以确保故事具有叙事的品质。叙事模型将故事模型作为目标受众个性化叙事的输入模型，也是数据故事化过程中的最终模型。

表 4 叙述类型与叙事模型

叙述类型	叙事模型	形式	内容
叙述者驱动型叙述	马提尼酒杯模型	是最常见的故事模型，杯子的长颈代表创作者的单方面叙事，宽阔的颈部代表读者参与的互动。	这个过程中，创作者首先根据数据分析建立一个完整的数据故事，然后由讲故事的人将建立的故事呈现给观众，传达创作者的观点，并为观众提供一个观点引发讨论，让听众开始讨论现有的数据故事并自由提问以便探讨。
	SUCCESs 模型	遵循 SUCCESs 原则（简单、惊奇、具体、信任和情感）。	简单是指把重点放在真正需要和想要传达的最重要的信息上，并保持短小精悍，减少理解和记忆的难度。惊奇是指打乱观众的理解，使情节或内容出人意料、令人惊讶，以吸引观众的注意力。具体是指提供具体的例子或具体的细节，使故事更清晰，更容易理解。可信指的是需要用可靠的来源来支持叙述者的说法，并使用公开的数据和外部证实来使故事更加可信。情感是指娴熟地使用一系列的情感，如快乐、悲伤、恐惧、惊讶、愤怒、担忧、爱和恨等

			来丰富故事本身，增加故事的吸引力，以引起观众的共鸣。
受众驱动型叙述	互动演示幻灯片模型	是一种让创作者和观众参与到故事讲述过程中的方式。	这种方法可以让听众更好地理解创作者所传达的意图，如果有必要，可以回顾数据叙述的每个阶段，以便更好地理解复杂的数据集和数据叙述的创建过程。
	下钻式叙事模型	以受众为中心，呈现出数页的数据可视化结果，让受众自己决定他们想看哪些细节和故事趋势。	在这种讲故事的模式中，创作者让受众选择他们想深入了解的故事，而受众则决定他们想了解的故事。

3.3.2 受众反馈

孙智中[20]等对数据故事化的评价与改进进行了研究，通过分析本文认为数据故事化的结果应该与业务需求进行比较和分析，受众应该能够根据自己的看法对数据故事进行评价，这样他们就可以与创作者进行互动，并提出进一步持续改进的建议。数据故事化不是一个一次性的活动，而是一个循序渐进的过程，涉及规划、实施、评价和改进等关键行动。

4 数据故事化的模型研究

在数据故事化的研究过程中，国内学者也提出了一些可用模型，包括有理论模型、应用模型与开发模型等。数据故事化的理论模型可以帮助组织厘清复杂或模糊的概念，梳理出相关基础理论的内容，有利于研究者了解其研究内容，还有利于数据利益相关者从中获取信息，进而做出正确的决策；应用模型为数据故事创作者的创作过程提供了模型化的参考，同时为数据故事化的实现提供了完整的理论路径；开发模型为数据故事的自动生成提供了方法，并为数据故事产品的开发提供了软件工程指导，可以推动数据故事工具的差异化发展及产业生态的培育。本文着重分析了其中几个具有代表性的模型（见表5）。

表 5 我国数据故事化模型比较

模型	提出者	组成	功能	异同与适用
数据故事视角下用户认知图式联结模型	王萍[19]Error: Reference source not found	模型由八种用户认知图式与认知图式的联结过程、反应及结果组成。	有利于研究者了解用户数据认知的需求，并为数据故事的采纳提供了理论基础。	与其他模型不同，它只适用于数据故事化的开端，用来了解用户的认知需求。

数据故事的金字塔模型	朝乐门[26]	模型从下到上由业务、数据、分析、洞察、故事模型、故事描述、故事叙述以及受众行为组成。	为数据故事创作者的创作过程提供了模型化的参考。	金字塔模型与实践路径理论模型都适用于数据故事化的整个创作过程，但其组成内容并不相同。
数据故事化实践路径的理论模型	周霞[21]	模型从叙事声音、多源数据、视觉效果、叙事逻辑、关联应用以及辅助信息 6 个维度构建组成。	为数据故事化的实现提供了完整的理论路径。	
基于结果的数据故事化评价模型	孙智中[20]	模型主要由外部变量、感知变量以及意向变量三大部分组成。	为数据故事化的效果评价提供了方向角度指导，有利于对数据故事化进行改进。	与其他模型不同，它只适用于数据故事化的结尾，用评价数据故事化的效果。
数据故事的工程化开发的参考模型	朝乐门[26]	模型由数据层、分析洞察层、故事模型层、故事脚本层、故事叙事层以及用户层六个层级组成。	为数据故事的自动生成提供了方法，并为数据故事产品的开发提供了软件工程指导，可以推动数据故事工具的差异化发展及产业生态的培育。	与其他模型不同，它是在金字塔模型的基础上构建的数据故事的分层实现和组件化研发的参考模型，适用于数据故事产品的工程化开发。

5 数据故事化的应用研究

数据存在于各行各业，包括政府、商业和科学研究，而数据故事化会使研究成功，并有助于发现和传播各种领域的专业知识。如何使用数据，讲述什么样的故事，是数据故事化会发挥作用的原因，并有助于其纵向和横向扩展。

本文通过总结我国现有的 21 篇文献中众多学者提出的有关数据故事化的应用领域，认为数据故事化的实践应用主要包括社交网络与自媒体、数据新闻传播、商业营销推广、档案内容挖掘、认知学习[26]（见表 6）。

表 6 数据故事化实践应用

实践应用	应用内容	应用示例
社交媒体网络	在开放、自主和多元的叙事系统中，自我实现的媒体内容生产者自发地跟踪当前的社会事件，解释数据，收集、组合和传播叙事内容，从观众的互动反馈中提取数据并产生共鸣和讨论，创造新的叙事视角和可视化的数据产品。	疫情情况展示[16]、文博展示、影视创作、社会记忆[25]

数据新闻传播	新闻业中的数据故事可以概括为如何获取已发布和未发布的数据，编辑如何分析数据，以及如何将数据可视化应用于新闻报道。	疫情数据传播[16]、政府开放数据[19]、数据新闻产品开发[1121]
商业营销推广	在商业领域扩大数据故事的应用，可以帮助组织和个人根据数据的可靠信息做出决策，并利用数字媒体增加营销活动的吸引力和流量。	商务演示[21]、市场营销[25]、数据故事产品开发、数据故事化专用工具平台[27]
档案内容挖掘	结合虚拟现实和全息投影等最新技术进行数据采集和展示虚拟、沉浸、互动和参与式的讲故事形式可以帮助观众更好地理解档案和艺术品的深层人文意义和历史，能更好地理解。	红色档案资源内容挖掘[14]、古籍内容挖掘、博物馆书籍内容挖掘[22]
认知学习	在与人脑和认知行为密切相关的领域，如教育和娱乐，数据讲故事在引导人类交流和参与方面有很大的潜力，是帮助解决复杂认知问题的重要工具。	解决数据科学“最后一英里”问题[13]、学术交流[18]

6 结论与展望

本文阐述了我国目前有关数据故事化的研究进展，从概念内涵、基本流程、相关模型以及应用实践 4 个方面对我国数据故事化的研究进展进行梳理和分析。研究发现，我国现阶段对于数据故事化的理论探索有了一定成果，但仍然存在着一些发展瓶颈：①我国有关数据故事化的研究主要以跟踪国外的研究为主，缺少一些原创性的研究；②目前的研究主要集中在理论分析以及实践应用等方面的模块化分析，缺少一套完整的数据故事化理论体系；③国内学者提出的现有模型主要集中在理论模型上，缺少面向不同业务需求的数据故事化模型的设计与创新；④目前的数据故事化还是主要依赖于可视化工具，缺乏开发专门针对支持数据故事化的开源工具；⑤目前的研究只提出了数据故事化可以应用的领域，但缺乏研究数据故事化在某个领域的具体应用。

为解决上述的发展瓶颈，未来的研究方向可以注重以下内容发展：①完善数据故事化的理论体系，进行系统性的研究，建立完整的知识体系，包括数据故事化的概念、理论、方法、技术、工具和实践；②优化数据故事化的算法，探索数据故事化的新模型，开发适用于不用实践场景的算法和模型以及自动调参的方法；③开发专门针对于数据故事化受众群体的用户研究框架，更准确地了解受众群体的需求，以便设计出更具效果的数据故事产品。④在当前发展的基础上，开发多样化的、易于使用的、富含故事和情感的、专门用于数据故事化的方法和工具。⑤在相关领域扩大数据故事化的应用范围，如政府开放数据

故事化与科研数据故事化等，并探索数据讲故事的具体应用，以提高该数据的可理解性、逻辑性、灵活性、趣味性和可读性，供公众再使用。

参考文献

[1] 周霞,王萍,张韫麒,宋婧馨,陈为东. 数据故事化实践应用分析——以数据新闻为例[J]. 图书情报工作,2021,65(14):119-127.

[2] Dykes B. Effective data storytelling: how to drive change with data, narrative and visuals[M]. John Wiley & Sons, 2019.

[3] Lee B, Riche N H, Isenberg P, et al. More than telling a story: Transforming data into visually shared stories[J]. IEEE computer graphics and applications, 2015, 35(5): 84-90.

[4] McDowell K. Storytelling wisdom: Story, information, and DIKW[J]. Journal of the Association for Information Science and Technology, 2021, 72(10): 1223-1233.

[5] Brolcháin N Ó, Porwol L, Ojo A, et al. Extending open data platforms with storytelling features[C]//Proceedings of the 18th Annual international conference on digital government research. 2017: 48-53.

[6] Ojo A, Heravi B. Patterns in award winning data storytelling: Story types, enabling tools and competences[J]. Digital journalism, 2018, 6(6): 693-718.

[7] Botsis T, Fairman J E, Moran M B, et al. Visual storytelling enhances knowledge dissemination in biomedical science[J]. Journal of biomedical informatics, 2020, 107: 103458.

[8] 张晨. 基于用户交互行为分析的数据故事建模方法研究[J]. 情报理论与实践,2021,44(02):185-191.

[9] 张晨,朝乐门,孙智中. 数据故事叙述的关键技术研究[J]. 情报资料工作,2021,42(02):73-80.

[10]蒋诗维. 时序性数据故事的叙事研究[D].太原:太原理工大学,2022.

[11]孙智中. 数据故事：概念及特征初探[J]. 情报科学,2022,40(09):167-175.

[12]肖纪文. 面向局部可解释性机器学习的数据故事生成方法研究[J]. 图书情报工作,2023,67,(2):98-107.

[13]朝乐门,张晨. 数据故事化:从数据感知到数据认知[J]. 中国图书馆学报,2019,45(05):61-78.

[14]赵红颖,张卫东. 数字人文视角下的红色档案资源组织：数据化、情境化与故事化[J]. 档案与建设,2021,(07):33-36.
[15]李成熙,文庭孝. 厚数据研究综述[J]. 高校图书馆工作,2022,42(01):8-14,28.
[16]侯雪林,应峻,宋士杰. 活动理论视角下疫情数据故事化的结构要素和过程探析[J]. 情报理论与实践,2022,45(08):68-75.
[17]陈昱彤,丁家友. 数据叙事的运行模型与关键问题[J]. 图书馆论坛:1-11.[2023-03-23].http://kns.cnki.net/kcms/detail/44.1306.G2.20220609.2106.009.html.
[18]孟刚. 数据故事化中的数据关键特征表达有效性研究[J]. 情报理论与实践,2023,46(02):117-126.
[19]王萍,周霞. 政府开放数据故事化实现模式研究[J]. 文献与数据学报,2020,2(02):3-13.
[20]孙智中,朝乐门,王锐. 数据故事化的评价与改进[J]. 情报资料工作,2021,42(02):81-89.
[21]周霞,王萍,张韫麒,宋婧馨,陈为东. 数据故事化实践应用分析——以数据新闻为例[J]. 图书情报工作,2021,65(14):119-127.
[22]张斌,李子林. 图档博机构"数字叙事驱动型"馆藏利用模型[J]. 图书馆论坛,2021,41(05):30-39.
[23]王海智. 可视化叙事在数据新闻中的应用研究[J]. 传媒,2021,(01):39-41.
[24]丁家友,唐馨雨. 数字人文视角下的数据叙事及其应用研究[J]. 情报理论与实践,2022,45(02):121-128.
[25]李锦绣,林泽斐. 面向数字人文的沉浸式数字叙事：现状、路径与展望[J]. 数字图书馆论坛,2022,(10):41-48.
[26]朝乐门. 数据故事的自动生成与工程化研发[J]. 情报资料工作,2021,42(02):53-62.
[27]易旎,朝乐门,张晨. 可视故事化：特征、方法与应用[J]. 情报资料工作,2021,42(02):63-72.

作者贡献说明：刘桂锋：整体构思，论文修改；吴雅琪：论文写作；刘琼：文献收集，论文修改。

Research progress in data storytelling in China: connotation, process, model, and application

Liu Guifeng Wu Yaqi Liu Qiong

Institute of Science and Technology Information, Jiangsu University

Abstract： [Purpose/Significance] Data storytelling is a more efficient, comprehensive, and updated way to present data analysis results. By analyzing representative literature on data storytelling research in China, this paper systematically sorts out existing research results, and further prospects research trends, providing reference and inspiration for future research. [Method/Process] The content analysis method is mainly used to summarize the content of 21 existing literature related to data storytelling research in China. The research progress of data storytelling in China is summarized and analyzed from four aspects: conceptual connotation, basic process, relevant models, and application practice. [Result/Conclusion] The research on data storytelling is still in its infancy in China, and there is significant room for improvement in theoretical research and practical applications such as content, processes, models, methods, and tools. Future research will focus on building an independent knowledge system and application practice system for data storytelling research in China.

Keywords： Data narrative; Storing data; Data science; Scientific data; data management